

Looking into the performance-based financing black box: Evidence from an impact evaluation in the health sector in Cameroon*

Damien de Walque, PhD. Development Research Group, The World Bank
Paul Jacob Robyn, PhD. Health Nutrition and Population, The World Bank
Hamadou Saidou, PhD. Health Nutrition and Population, The World Bank and University of Paris Dauphine
Gaston Sorgho, DrPH. Health Nutrition and Population, The World Bank
Maria Steenland, SD. Population Studies and Training Center, Brown University

Key messages:

Several studies have shown a positive impact of performance based-financing (PBF) on health service coverage, often coupled with improvements in quality, but relatively little is known about the mechanisms driving those results. This paper presents results of an impact evaluation in Cameroon designed to isolate the role of specific components of the PBF approach: explicit financial incentives linked to results, additional resources available for health providers not linked to performance, and enhanced supervision, coaching and monitoring.

Overall, results indicate that, when compared with the pure control group, PBF in Cameroon led to significant increases in utilization for several services (child and maternal vaccinations, use of modern family planning), but not for others like antenatal care visits and facility-based deliveries. In terms of quality, PBF increased the availability of inputs and equipment, qualified health workers, led to a reduction in formal and informal user fees, but did not affect the content of care.

However, for many positively impacted outcomes, the differences between the PBF group and the group receiving additional financing not linked to performance are not significant, suggesting that additional funding rather than the explicit incentives might be driving improvements.

Keywords: Performance-based Financing; Health; Africa

JEL classification: I15 J13; O15.

Word count: 5998

Abbreviated running title: Performance-based financing for health in Cameroon

* We are extremely grateful for a fruitful collaboration with the Ministry of Health of Cameroon and in particular with his Excellency Andre Mama Fouda, Minister, Enandjoum Bwanga, Emmanuel Maina Djoulde, Victor Ndiforchu, and all the Regional Health Delegates, Performance Purchasing Agency Managers and staff and from the North-West, South-West and East Regions. The baseline and endline data were collected by the Institut de Formation et de Recherche Démographiques (IFORD) with special thanks to Gervais Beninguisse, Didier Nganawara, Evina Akam and all staff implementing the baseline and endline surveys. This impact evaluation is funded by the Health Results Innovation Trust Funds (HRITF) at the World Bank. We are grateful to Chantelle Boudreaux, Emanuela Di Gropello, Guenter Fink, Gyuri Fritsche, Elisabeth Huybens, Gil Shapira, Jean Claude Taptue Fotso, Robert Soeters and Omer Zhang for useful inputs, comments and discussions. Ethical clearance was obtained for the impact evaluation and both baseline and household surveys from Cameroon's "Comité National d'Ethique/National Ethics Committee" (authorization 289/CNE/SE/2011 on December 19, 2011 for the baseline survey and 2015/02/555/CE/CNERSH/SP on February 16, 2015 for the endline survey).

The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

Corresponding author: Damien de Walque, The World Bank, Development Research Group, 1818 H. Street NW, Washington, DC 20433, USA. ddewalque@worldbank.org +12024732517. <http://orcid.org/0000-0003-1592-7602>

Looking into the performance-based financing black box: Evidence from an impact evaluation in the health sector in Cameroon

Abstract

Performance-based financing (PBF) is a complex health systems intervention aimed at improving coverage and quality of care. Several studies have shown a positive impact of PBF on health service coverage, often coupled with improvements in quality, but relatively little is known about the mechanisms driving those results. This paper presents results of a randomized impact evaluation in Cameroon designed to isolate the role of specific components of the PBF approach with four study groups: i) PBF with explicit financial incentives linked to results, ii) direct financing with additional resources available for health providers not linked to performance, iii) enhanced supervision and monitoring without additional resources and iv) a control group. Overall, results indicate that, when compared with the pure control group, PBF in Cameroon led to significant increases in utilization for several services (child and maternal vaccinations, use of modern family planning), but not for others like antenatal care visits and facility-based deliveries. In terms of quality, PBF increased the availability of inputs and equipment, qualified health workers, led to a reduction in formal and informal user fees, but did not affect the content of care. However, for many positively impacted outcomes, the differences between the PBF group and the group receiving additional financing not linked to performance are not significant, suggesting that additional funding rather than the explicit incentives might be driving improvements. In contrast, the intervention group offering enhanced supervision, coaching and monitoring without additional funding did not experience significant impacts compared to the control group.

1. Introduction

Confronted with slow progress on the health-related Millennium Development Goals (United Nations Statistical Division, 2015; Wagstaff, Bredekamp and Buisman, 2011) and with evidence that productivity and quality are low in health facilities in low and middle-income countries (Das, Hammer and Leonard 2008), in the last decade various countries experimented with Performance Based Financing (PBF) in the health sector. PBF is a broad health system reform using supply-side incentives to increase coverage and quality of essential health services, as well as efficiency and equity, often with a focus on maternal and child health. PBF program models differ but all involve the purchasing of health services using a pre-defined list of services and prices (Fritsche, Soeters, and Meessen, 2014). Many PBF programs also involve increasing health facility autonomy.

PBF has been implemented in 51 countries, 25 of them in Africa (Vellez, 2019). Several studies have shown a positive impact of PBF on health service coverage, often coupled with improvements in quality. An early impact evaluation in Rwanda where districts were randomly assigned to treatment (PBF) and comparison (input financing with matched financial resources) found large and statistically significant positive impacts on institutional deliveries and preventive care visits from young children and also on quality of prenatal care (Basinga et al., 2011). After this promising start, many other quasi-experimental studies have shown similarly positive results (Gertler, Giovagnoli and Martinez, 2014; Ir et al., 2015; Zeng, Cros, Wright, & Shepard, 2013) while several others have shown favorable results for many – though not all – outcomes assessed (Binyaruka et al., 2015; Bonfrer, Soeters, et al., 2014; Bonfrer, Van de Poel, and Van Doorslaer, 2014; Falisse, Ndayishimiye, Kamenyero, and Bossuyt, 2015; Ahmed, Arur, de Walque and Shapira 2019). Despite this, other studies have found only limited positive results and the research community has not reached a consensus about the effectiveness of PBF at increasing health service coverage (Paul, Albert, Bisala et al. 2018; Huillery and Seban, 2019).

While the evidence about the impacts of PBF accumulates, few studies have examined the factors and mechanisms that influence its impact, an area of substantial theoretical and practical significance since PBF often involves a package of interventions: linking payment and results, independent verification of results, managerial autonomy to facilities and enhanced systematic supervision and coaching of facilities. Shen et al. 2017 and a forthcoming study in Nigeria (Kandpal et al.) examine the difference between PBF and additional financing while Shapira et al. 2018 compare incentives for community health workers to in-kind demand side incentives in Rwanda. In this study, we designed the impact evaluation of the PBF package in Cameroon to try to understand better the role of some of these mechanisms. In particular, we tried to isolate the role of explicit financial incentives as opposed to additional funding not linked to performance, as well as separating the impact of enhanced supervision and monitoring. We did so by randomizing facilities to four arms including i) the full PBF intervention linking payments to improvements in the quantity and quality of health services, ii) direct financing with additional funding not tied to performance, iii) additional supervision and monitoring without additional funds and iv) a control group where business-as-usual was conducted. We evaluated the impact of those interventions focusing on the utilization of maternal and child health services, measures of quality of care (structural and content of care) and out-of-pocket payments by service users.

2. Context and intervention

Health system context in Cameroon

Cameroon did not achieve Millennium Development Goals 4 & 5 which called for large reductions in maternal and child mortality (MDGs). Child mortality declined in Cameroon by approximately 21% between 1991 and 2014 (ICF International, 2012; Institut National de la Statistique, 2015; Institut National de la Statistique (INS) et ICF International, 2012); nonetheless, approximately 1 in 10 children still die before their fifth birthday (Alkema et al.,

2015; The United Nations Children's Fund, 2015). Though many factors affect maternal mortality, one potential explanation for the country's ongoing high maternal mortality rate is the slow progress towards adequate prenatal care (59% in both 2004 and 2014) and facility delivery (62% in 2004 and 65% in 2014) in the decade leading before Cameroon's PBF program began.

Several aspects of the health care financing landscape in Cameroon contribute to the low quality of primary health care service provision, and sub-optimal coverage of essential maternal and child health care services. Cameroon is one of the wealthier countries in the Central African region, and the country's health spending of \$59 per capita in 2014 is relatively high (World Health Organization Global Health Expenditure database, 2016) (The World Bank, 2016). Cameroon spends \$10 dollars more than the average for Sub-Saharan Africa (excluding South Africa) (Bove, Basile., Robyn, & Singh, 2013); however, despite this relatively high level of overall spending, the share of per capita total health spending paid for by the government was only US \$14 (i.e. 22%). Much of the remaining health spending is paid for through out-of-pocket users fees (Ministère de la Sante Publique, 2016).

Further, the allocation of funding is favors central-level administration and tertiary care which limits the funding available at primary care facilities. Less than 10% of the budget is allocated to service providers and decentralized levels of the Ministry of Public Health nationwide at the regional, district and health facility level (The World Bank, 2017). Facilities receive financing from the government for recurrent and to some extent capital expenditures based on historical budgeting This has resulted in a scarcity of funds to meet operating expenses incurred in the day-to-day business of a district health system (e.g., consumables, drugs, regular maintenance, community outreach, etc.) (Ministère de la Sante Publique, 2016).

Performance-based financing intervention

The administrative and technical aspects of the PBF program in Cameroon were managed at the regional level by Performance Purchasing Agencies (PPA). PPAs are

autonomous entities that are contracted by the government to generate performance contracts with health facilities, verify health facility data, and manage funding intended for health care providers through PBF.

The PBF package of interventions included the following elements:

- Linking payment and results, including performance bonuses for health workers
- Independent monitoring of results
- Systematic supervision of health facilities defined as regular supervision by an external supervisor from the district hospital team using a structured checklist and providing immediate feedback to facility staff on problems identified and potential solutions to improve service delivery. It included monitoring compliance with national user fee guidelines
- Limited managerial autonomy to facilities, i.e. autonomy over use of resources combined with the ability to hire *additional* staff using health facility income and managerial discretion.

PBF contracts were signed for three months. Health facilities were responsible for completing registers with information on the number of patients seen for each health service and a monthly activity report/declaration form. This report and registers were used to document reported health service provision and were used as the primary basis for service verification. Technical Appendix A1 describes the service verification procedures. The PBF program incentivized curative care (e.g., hospitalizations and minor surgeries), preventive care (e.g. vaccination, infectious disease screenings) and reproductive health care (e.g., deliveries, family planning, antenatal care) services. Appendix Table A1 provides a complete list of health services subsidized by the program, and the subsidy amounts.

Following verification, performance payments were provided monthly. Payments included a fixed amount for each health service rendered scaled by a quality score and an equity bonus that advantaged poorer and more remote health districts. Quality assessments were conducted quarterly and based on a standardized checklist primarily measuring structural

quality of care (facility management, hygiene and sanitation) as well as, to a lesser extent, specific attributes of service delivery (correct management of cases). Technical Appendix A2 provides a detailed description of the calculation of performance payments.

After finalizing payments to PBF facilities, the direct financing group received a fixed per capita budgetary supplement (unrelated to performance) matching the per capita budgetary allocation for the PBF facilities. Facilities in the direct financing group were provided the same supervision and monitoring and managerial autonomy over the budgetary supplement received. The facility managers in both the PBF and the direct financing groups had the autonomy to hire staff with their PBF revenues or budgetary supplement received, and to fire these staff if necessary. Facilities in the enhanced supervision and monitoring group received no additional resources but the same supervision and monitoring as in the PBF and direct financing facilities. District-level supervisors responsible for monitoring and supervision used the same tools and received the same supplementary payments for visits to facilities in all three treatment groups.

The mean PBF and direct financing subsidies received by the facilities were between \$342 and \$476 per quarter per health worker, on average more than doubling the existing baseline funding from the government (estimated between of \$319 and \$425 per quarter per health worker (see Appendix table A8). In the two other groups (enhanced supervision and monitoring) facilities obtained financing from the government based on historical budgeting without additional resources or inputs. National user fee caps, and facility user fee rates, were published on a signboard placed in all study group health facilities. Monitoring of adherence to national guidelines was included as part of the monitoring and supervision intervention in the PBF, direct financing and enhanced supervision and monitoring facilities. A summary table describing the intervention groups is provided in Appendix Table A2.

The Cameroon PBF program evaluated in this paper began in July 2012 in the North West and South West regions, with four districts included in each region. In October 2012 the program expanded to the East region, covering all 14 health districts in the region, while another

pilot had been separately run in 4 districts in the Littoral region. Out of the 26 health districts throughout Cameroon implementing PBF, 14 districts were included in the impact evaluation (see Figure 1). The other 12 (4 in Littoral and 8 in the East) had already begun implementing some form of PBF before the impact evaluation baseline survey was conducted or were added after it was conducted.

3. Evaluation Design and Methods

Table 1 describes the four study groups formed by randomizing “Centres Médicaux d’Arrondissement” (CMAs), or arrondissement primary health centers with a medical doctor on staff, and Centres de Santé Intégrés (CSIs) (primary health care centers without a doctor). The randomization for this study was at the health facility level and was stratified by region, urban/rural location and facility type (public/private). Public randomization ceremonies attended by the health facility management were held in each region between February and June 2012, just prior to the launching of the PBF program in each region (De Walque, Robyn, and Sorgho, 2013).

The number, type, and percent private of study health facilities in each study district are shown in Appendix Table A3. All public and private health facilities in the 14 study districts that were officially registered with the Ministry of Public Health were eligible for inclusion in the study¹.

Data sources

¹ All district hospitals in 14 health districts were included in the full PBF arm, because district hospitals play a critical role in supervising and acting as source of referral services for all facilities in the district. District hospitals did supervise, and support treatment and comparison group CMAs and CSIs differently based on the group they are assigned to. However, data from these hospitals are not used for making inferences about the impact of PBF nor included in the analyses reported here.

The evaluation relied on two main sources of data to answer the impact evaluation research questions identified:

Household surveys: A household survey implemented at baseline in 2012 (i.e., before implementation of PBF began), and at endline in 2015 (i.e., after PBF was implemented for more than two years). In both rounds, the primary inclusion criteria for the household survey was that household must have included at least one woman who had been pregnant in the 24 months preceding the survey. A village household listing exercise was first conducted to identify all village households. At baseline 16 of the households identified in the listing exercise were randomly selected to be surveyed in each village. The survey team attempted to revisit all baseline households at endline. However, many baseline households could not be located or were no longer eligible at endline. When this occurred, baseline households were replaced using the nearest neighbor as recorded in the listing exercise. An additional four households were added to the household sampling roster at endline such that a total of 20 households were sampled in each village for the endline survey.

Facility-based surveys: A facility-based survey was also implemented at baseline and at endline. Facility level provision of health services in the six months before the survey was assessed using patient registers. A survey team member observed consultations with a systematic random sample of patients under five presenting with a new condition (i.e., not for follow-up visits or routine) and new antenatal care (ANC) clients. As primary care facilities do not offer ANC services on all days of the week – typically these are offered 2 days each week – the ANC module was not conducted at all health facilities. During the baseline survey, 5 under-5 and 5 ANC observations were conducted at each facility where these modules are implemented. After finding that many health facilities did not offer ANC on the day of the survey at baseline, during the endline survey enumerators were asked to interview as many women receiving ANC on the day of the survey as possible to increase the sample size. Technical Appendix A3

provides additional information on the procedures used during direct observation of ANC and child health consultations.

Study outcomes:

We selected a set of PBF incentivized and non-incentivized study outcomes based on the following criteria: information reported and easily available in the health facility registers, procedures frequent enough to be reported in most facilities, outcomes for which information can be compared in the facility and household surveys. For brevity, we focus this study on maternal and child health outcomes and therefore did not include the HIV and TB indicators

Primary outcomes: Outcomes measured using facility register data recording health service provision during the six months before the survey except when otherwise noted

- Facility delivery
- Antenatal care visits
- Receipt of tetanus vaccine during pregnancy
- Postnatal care
- Modern contraception: contraceptive implants, injectables, oral contraceptive pills, and intrauterine device
- Third dose of polio, meningitis, and measles vaccines
- Quality of antenatal care/child health consultations: Measured using a quality score calculated using the share of World Health Organization recommended ANC/ child health consultation service components completed during observed consultations².

Outcomes measured using self-reported household survey data:

Among women who had been pregnant in the previous 24 months:

² The measures for quality of care of ANC and child health consultations are more extensive than the “correct management of cases” in the PBF quality bonus formula which does not include direct observations of clinical interactions.

- Skilled delivery: Most recent childbirth in a health facility with a skilled attendant
- At least two ANC visits: At least two ANC visits during the most recent pregnancy
- Tetanus vaccine during pregnancy: Tetanus vaccine receipt during the most recent pregnancy
- Postnatal care: Postnatal care receipt during the most recent pregnancy
- Modern contraception: intrauterine device, injectables, implants, oral pills, diaphragm, foam/jelly & lactational amenorrhea among on-pregnant fertile women of reproductive age (15 – 49 years) with a current sexual partner
- Child vaccination: Receipt of oral polio vaccine, yellow fever, diphtheria and whooping cough (DTC), measles, and Bacillus Calmette–Guérin (BCG) among children between 12 – 23 months of age

Secondary outcomes:

- Health care spending: Household out-of-pocket spending for treatment of illness in the four weeks before the survey.
- Patient satisfaction with ANC: Exit interviews conducted with women attending ANC visits
- Availability of facility staff: Number of nurses present at health facility on the day of data collection
- Availability of equipment and supplies: Composite indicator of the availability of basic clinical equipment.³

We also examine several non-incentivized services (growth monitoring, bednet use, stunting, underweight and wasting among children) to test whether the study interventions led facilities to reduce facility resources towards non-incentivized care, resulting in an unintended decline in provision of these services.

³ The indicator included information on the presence of a clock, a child weighing scale, height measure, tape measure, adult weighing scale, blood pressure instrument, thermometer, stethoscope, fetoscope, otoscope, flashlight, stretcher, and wheelchair. Scores indicate the proportion of these thirteen pieces of equipment that was available in the facility. Point estimates indicate the estimated impact on this score.

Statistical Methods

Relying on random assignment at the facility level, we used difference-in-differences regression for the health facility and household data analysis. The regression specifications we used are detailed in appendix A4.

The analysis of the household survey in this report starts from the assumption that household members seek care in the health facility closest to where they live, or in other words that people living in the catchment area of a facility obtain health care in that facility. However, it is apparent from the baseline survey data that households do not always seek care from the closest health facility in their health zone. We recognize that this assumption is not always verified and that therefore the results might suffer from attenuation bias. We address this issue in further detail in technical Appendix A4 entitled Analysis of Household Care Seeking Behavior. Overall, we did not find significant evidence that health care seeking behavior was influenced by the introduction of PBF. In addition, the following other considerations motivated our choice to keep this “intent-to-treat analysis”: i) in many cases, the data collected about which health facility was visited allowed us to find out whether the visited facility was the closest one, but in case it was not, did not allow us to ascertain to which study group the visited facility pertained; ii) when we could ascertain to which study group the visited facility pertained, such bypassing behavior is clearly endogenous and assigning to the household bypassing its closest facility the study group of the facility actually visited would lead to endogeneity bias and iii) the results from the household survey analysis are broadly consistent with the results from the health facility survey analysis, which are not affected by the measurement error introduced by the health care shopping behavior of households.

We adopt two strategies to address the large number of outcomes that we examine to avoid overemphasizing any single significant result. First, we create indexes combining all the outcomes for each family of outcomes following Kling, Liebman, and Katz (2007). To construct

the indexes, we define each outcome so that higher values correspond with better outcomes. Then we standardize each outcome into a Z-score by subtracting the mean and dividing by the standard deviation of the older cohort born in low intensity regions. We then average all of the Z-scores and standardize the average relative to the control group. We then estimate the effect of the cash transfers on these standardized outcome indexes. Second, we correct for the potential issue of simultaneous inference using multiple hypothesis testing. Following Anderson (2008), we compute, and report sharpened False Discovery Rate (FDR) q-values. The FDR is the expected proportion of rejections that are type I errors (false rejections).

4. Results

Balance at baseline

The sample was balanced on nearly all of the individual and household and facility characteristics assessed. Means, t-tests between groups, and F-tests of overall differences are provided in Appendix Tables A4 to A7. The subsidies provided to the direct financing group were not statistically different from the payment amounts provided to the PBF treatment group (Appendix Table A8).

Facility survey utilization results

This section describes the results of PBF on health services provision as recorded in facility registers⁴. Relative to the control group there were no statistical differences in mean monthly provision of skilled delivery in the full PBF and the enhanced supervision and monitoring group; however, there was a relative increase of approximately 2 deliveries per month in the direct financing group (direct financing 1.855, p-value = 0.071), even though that result was not robust to the multiple hypothesis testing adjustment (q-value 0.166) (Table 2).

⁴ To assess the reliability of these data, we examined health service counter-verification data that was collected routinely as part of the PBF program design through community client satisfaction surveys and is further detailed in Appendix A1.

Relative to the control group there was an increase in antenatal care in all three treatment groups; however, none of these differences were statistically significant. Compared to the change observed in the control group, there was a positive and statistically significant difference in the PBF and direct financing groups for the provision of tetanus toxoid vaccine during pregnancy (column 3), postnatal care (column 4)⁵ and modern contraception (including contraceptive implants, injectables, oral contraceptive pills, and the intrauterine device, see column 5). Though the interaction term was positive, there was no statistical difference for the additional supervision group (EMS).

There was an increase of 4.583 final polio vaccinations relative to the control group in the PBF group (column 6), and this difference was statistically significant ($p=0.035$) but not robust to the multiple hypothesis testing adjustment ($q\text{-value}=0.125$). The coefficient on the interaction term was also positive in the direct financing and enhanced supervision and monitoring groups but the difference was not statistically significant. For meningitis vaccination provision (column 7), of the three treatment groups, only the change in the direct financing group was statistically different from the control group (21.931, $p\text{-value}=0.050$) but not robust to the multiple hypothesis testing adjustment ($q\text{-value}=0.158$). Finally, for the measles vaccine, there was no difference between the change observed in the control group, and the changes in the treatment groups (Table 2, column 8).

In column 9, we regroup the outcomes related to pregnancy and maternal health and find that both PBF and direct financing improved utilization compared to the control group, but that their impacts were not different from each other. In column 10, we regroup the child immunization indicators and find that PBF improved vaccinations compared to the control group ($p= 0.057$; $q\text{-value} = 0.11$). Finally, in column 11, we regroup all 8 maternal and child health utilization indicators that both PBF and direct financing improved utilization compared to

⁵ For the PBF group, the increase in postnatal care was not robust to adjusting for multiple hypothesis testing ($q\text{-value} = 0.163$).

the control group and that the PBF impact was larger but not significantly so than the direct financing ($p=0.145$). However, if we exclude postnatal care from the index (column 12), both PBF and direct financing again improved utilization and the PBF coefficient is significantly larger ($p=0.091$).

Household survey results

No statistically significant treatment effects were observed for the PBF and the additional financing groups for study outcomes related to care during pregnancy in the household survey data⁶ (Table 3). However, skilled delivery, attendance to at least two antenatal care visits and postnatal care declined in the group receiving only improved supervision relative to the control group, but these results are not robust to the adjustment for multiple hypothesis testing. Results from testing the equality of coefficients show that for skilled delivery the direct financing intervention outperformed the PBF group. For antenatal care, facilities with the PBF intervention performed better than those facilities with only enhanced supervision and monitoring. We regrouped the reproductive health services in two indices, one in column 6 including delivery, ANC and tetanus vaccination during pregnancy and not including postnatal care, while the index in column 7 adds postnatal care. No impact of PBF or DF is found on these indices of utilization indicators collected at the household level. We did not include modern family planning in the indices because the question was asked of a different sample of women, but since no impacts of PBF or direct financing were found on family planning in column 5, this is unlikely to make a difference.

⁶ The strong effect of PBF and direct financing on family planning in the facility register data differs from the null finding we found in the household data (Table 3, column 5). The difference in the findings between data sources could be explained by several factors. One possibility is that since the household survey oversampled recently pregnant women (as this was an inclusion criteria), the need for modern family planning is less important among women who have recently shown a willingness to have children. Moreover, there is evidence that women might have been uncomfortable disclosing their use of family planning in the household survey. For example, 47% of women report that their husbands are against using family planning to avoid pregnancy.

For immunization, Table 4 focuses on vaccination coverage, growth monitoring, bednet use and anthropometrics among children. Columns 1 and 2 show that for immunization coverage the PBF group outperformed the three other study groups, irrespective of the way the dependent variable was created⁷. These impacts however are not robust to the adjustment for multiple hypothesis testing. The remaining outcomes in table 4 were non-incentivized by the PBF scheme⁸. Growth monitoring increased by approximately 3 percentage points more in the group that received direct financing (0.031, p-value = 0.071, but was not robust to the multiple hypothesis testing adjustment) than the control group. Otherwise, no statistically significant effects were measured, as confirmed in the index in column 8.

Health care spending

Neither the change in the control group on spending for official provider fees, nor the differences between the treatment groups and the control group were statistically significant (Table 5, column 1). There was an increase of approximately 2052 CFA francs (\$3.31) in the amount of unofficial provider fees paid in the control group between 2012 and 2015. The difference between the PBF group and the control group over the study period was a decrease of 2254 CFA francs (\$3.64), and this difference was statistically significant, but not robust to the multiple hypothesis testing adjustment. Relative to the control group, unofficial payments also

⁷ The facility results in Table 2 documented large declines in vaccine provision. This is not necessarily inconsistent with the household vaccine coverage results, which showed large increases over time, and a particularly large change in the PBF group. According to the recommended vaccine schedule, most childhood vaccines should be received between birth and 12 months of age. Three of the five vaccines we assessed (polio, BCG, and DTC) are completed within 8 weeks after birth. Yellow Fever and Measles should be done at 9 months. Technically there is a second dose for measles, at least in the [2019 WHO recommendations](#), but we only assess one dose in the surveys. To avoid including infants who, due to their age, should not yet have finished their vaccine schedule, household level coverage only included children between 12 – 23 months old. Consequently, the children included in the household data would have received vaccinations prior to the months covered in the facility level data at endline (i.e. before the vaccine stock-out). Both sets of results show that the PBF group, and to a lesser degree, the direct financing group, performed much better than the control group. Therefore, the findings from the household and facility data showing large and statistically significant differences between groups are consistent, and we believe that the level changes can be explained by a country level shock affecting vaccine availability that affected the facility, but not the household data.

⁸ Those outcomes, including height and weight measures were assessed among children under five years of age; however, children under 12 months were not included in the growth monitoring analysis.

declined in the direct financing group and the enhanced supervision and monitoring group; however, these differences were not statistically significant (Table 5, column 2). Compared to the control group, laboratory and x-rays fees declined by 1473.44 CFA francs (\$2.38) in the PBF group, and this difference was statistically significant (p-value = 0.060), but not robust to the multiple hypothesis testing adjustment (Table 5, column 3). Finally, for transportation fees there was a statistically significant (but not robust to the multiple hypothesis testing adjustment) difference between the control group, and the direct financing group with a decrease of 495.14 CFA francs or \$0.80 (Table 5, column 4). In column 5, the index regrouping the four previous outcomes indicates that PBF led to a statistically significant decline in patient spending compared to the control group.

Quality of care

As shown in Table 6 column 1, there is no evidence of impact in any of the three intervention groups on the quality of child health consultations. are presented in Table 6 column 2⁹. The quality of antenatal care consultations increased during the study period by 12.9 percentage points between baseline and endline (p<0.05) in the control group. However, there were no differences in any of the treatment groups in the change in ANC quality relative to the full control.

Structural Quality

The increase in the number of nurses in the PBF group was greater compared to the full control group (p-value = 0.01) (Table 7, column 1). Adding the coefficient on the interaction term of PBF and post to the coefficient on the post indicator (0.191+1.222=1.413) indicates that there was an average increase of almost 1.5 nurses present in PBF facilities over the study

⁹ We are grateful to Chantelle Boudreaux who conducted that analysis.

period. Both the PBF and the direct financing interventions resulted in large and statistically significant improvements in the availability of basic clinical equipment (column 2) and delivery equipment (column 4). There was no measurable difference in the impacts of the PBF and direct financing arms, but there was statistically significant difference between the PBF intervention and the supervision and monitoring arm (Table 7, columns 2 and 4). Those results are robust to the multiple hypothesis testing adjustment and are also present in the index regrouping the equipment and drug availability outcomes in column 9.

Patient satisfaction

Relative to the pure control, the PBF group was associated with an 8.6 percentage point increase in satisfaction for the ANC visits ($p = 0.077$) (Appendix table 9, column 1) and 9.9-percentage point increase in satisfaction ($p < 0.05$) for visits with children under age 5 (Appendix table 10, column 1).

Discussion and Conclusions

Overall, the impact evaluation results reveal significant increases in the PBF arm for several indicators (child and mother vaccinations, use of modern family planning), but not for others such as ANC visits and in-facility deliveries. Structural quality as measured by equipment availability, staff presence and staff satisfaction, improved in the PBF group. The majority of these impacts are confirmed when indices for family of outcomes are considered and are robust to adjustment for multiple hypothesis testing. However, despite an increase in providers and supplies available at health facilities, PBF did not increase the completeness of service provision during ANC and child health consultations. Importantly, out-of-pocket health expenditures decreased for household in the PBF arm and this decrease in revenue did not come at the cost of process quality: there were no negative effects on completeness of services and advice provided

during antenatal visits and consultations for children under 5. Client satisfaction also increased for medical consultations for children.

While most the improvements measured for PBF were also observed in the direct financing arm (or cannot be distinguished statistically from each other, especially after adjusting for multiple hypothesis testing), few improvements were observed in the group offering only enhanced supervision and monitoring, suggesting that this intervention was not sufficient¹⁰. The difference between the PBF and the direct financing is subtle because the two interventions share many similarities: the same supervision and monitoring mechanisms, the same level of managerial autonomy and increased financing. The only difference was that in the PBF group the additional financing was linked to the performance of the individual facility while in the direct financing group it was linked to the average performance of the PBF facilities in the same district. It is possible that this distinction might not have been salient enough among the health facility management and staff for them to act upon it and modify their practice, explaining overall similar results. Further, the absence of impacts of PBF and direct financing on non-incentivized indicators (see table 4), while reassuring because it rules out negative spillovers, doesn't allow unpacking differences in the mechanisms at play in both arms.

Study limitations

This study has several limitations. Randomization at the health facility level (as opposed to the district level) increased the likelihood that neighboring facilities from different groups might learn from each other or from supervising staff and apply principles outside their treatment group. Further, district leadership may not have consistently maintained randomization due to the complexity of the intervention design randomized within their district facilities. Though we find that bypassing was not affected by PBF (Appendix A4 and Appendix

¹⁰ We cannot however make predictions on what would have been the of impact additional financing without supervision and monitoring since all three intervention groups included the reinforced supervision and monitoring.

Tables A12-A15), it may have attenuated the results in analysis using household data. Finally, PBF in Cameroon targeted many health services, and facility operating procedures leading to a large number of potential study outcomes. We focus on maternal and child health outcomes, constructed indices regrouping family of outcomes, include q-values adjusting for multiple hypothesis testing.

In our analysis, the facility and household results often differ substantially: in the facility data, PBF and additional financing have a positive effect on many of the outcomes whereas in the household data PBF has a much more limited effect. Beyond the different data collection methods, measurement error, including error due to health care bypassing behavior as described above, might attenuate the estimates in the household data. Further, treatment facilities might have been more likely to report service provision. We do not expect that increased reporting in treatment facilities is due to fraud, as strong verification mechanisms – including random phone calls to reported patients – were in place. Instead, treatment facilities might record clinical activity more completely since they are incentivized to do so.

Our findings that PBF and direct financing have similar impacts stands in contrast with the results from the initial impact evaluation in Rwanda where districts were randomly assigned to a PBF group and a comparison with input financing with matched financial resources. This study found statistically significant positive impacts on institutional deliveries and preventive care visits from young children and also on quality of prenatal care (Basinga et al., 2011). However, more recent studies in Zambia (Shen et al. 2017) and a forthcoming study in Nigeria (Kandpal et al.) find more limited differences between PBF and additional financing, consistent with our findings.

While some previous literature has suggested that weak management and lack of oversight explain weak performance in the private sector, we find that enhanced supervision and monitoring alone are not sufficient to improve maternal and child health indicators (Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013). This finding suggests that increased financing is

an essential part of strengthening primary health care provision.. The role of effective management in determining the quality of public health service provision in low-income countries remains understudied. Measurement tools that can capture the effectiveness of management in public facilities, better documentation of which aspects of management are weakest, and a better understanding of which aspects of management matter most for service provision may be required before effective interventions to improve management can be developed.

Conclusion:

The results of this study suggest that the additional financing without verification of services may be as effective as standard performance-based financing models. Given the widespread implementation and costs of the standard performance-based financing model, future studies should focus on understanding which aspects of performance-based financing drive results. In particular, further studies, or efforts to pool the results of existing studies, to compare PBF and direct financing would provide important guidance to policy makers.

References

- Ahmed, Tashrik, Aneesa Arur, Damien de Walque and Gil Shapira. 2019. Incentivizing quantity and quality of care: Evidence from an impact evaluation of performance-based financing in the health sector in Tajikistan. World Bank Policy Research Working Paper 8951.
- Alkema, L., Broaddus, E., Chou, D., Hogan, D., Mathers, C., Moller, A.-B., . . . Zhang, S. (2015). Trends in Maternal Mortality: 1990 to 2015 Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division In W. H. Organization (Ed.): World Health Organization
- Anderson, Michael L. (2008) Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects, *Journal of the American Statistical Association*, 103:484, 1481-1495.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. Soucat, J. Sturdy and C. M. Vermeersch (2011). "Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation." *Lancet* 377(9775): 1421-1428.
- Binyaruka, P., Patouillard, E., Powell-Jackson, T., Greco, G., Maestad, O., & Borghi, J. (2015). Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. *PLoS One*, 10(8), e0135013. doi:10.1371/journal.pone.0135013
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. *The Quarterly Journal of Economics*, 128(1), 1-51.

Bonfrer, I., Soeters, R., Van de Poel, E., Basenya, O., Longin, G., van de Looij, F., & van Doorslaer, E. (2014). Introduction of performance-based financing in Burundi was associated with improvements in care and quality. *Health Affairs*, 33(12), 2179-2187.

Bonfrer, I., Van de Poel, E., & Van Doorslaer, E. (2014). The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. *Social Science & Medicine*, 123, 96-104.

Bove, A., Basile., P., Robyn, P. J., & Singh, R. (2013). Towards Greater Equity: A Special Focus on Health. Cameroon Economic Update In The World Bank (Ed.). Washington DC: The World Bank.

Das, Jishnu, Jeffrey Hammer and Kenneth Leonard. 2008. "The Quality of Medical Advice in Low Income Countries". *Journal of Economic Perspectives*, 22(2): 93–114

de Walque, D., Robyn, P. J., & Sorgho, G. (2013). Public randomization ceremonies. *The World Bank Development Impact Blog*. Retrieved from <http://blogs.worldbank.org/impac evaluations/public-randomization-ceremonies>

- Falisse, J. B., Ndayishimiye, J., Kamenyero, V., & Bossuyt, M. (2015). Performance-based financing in the context of selective free health-care: an evaluation of its effects on the use of primary health-care services in Burundi using routine data. *Health Policy Planning, 30*(10), 1251-1260. doi:10.1093/heapol/czu132
- Fritsche, G. B., Soeters, R., & Meessen, B. (2014). *Performance-based financing toolkit*: World Bank Publications.
- Gertler, P. J., Giovagnoli, P. I., Martinez Sebastian (2014). Rewarding provider performance to enable a healthy start to life: evidence from Argentina's Plan Nacer. *World Bank Policy Research Working Paper* (6884).
- Huillery, E., & Seban, J. (2019). Financial Incentives, Efforts, and Performances in the Health Sector: Experimental Evidence from the Democratic Republic of Congo. Forthcoming *Economic Development and Cultural Change*.
- ICF International. (2012). The DHS Program STATcompiler. Retrieved June 10, 2015, from <http://www.statcompiler.com>
- Institut National de la Statistique. (2015). Enquête par grappes à indicateurs multiples (MICS5), 2014, Rapport de résultats clés. Yaoundé, Cameroun: Institut National de la Statistique.

Institut National de la Statistique (INS) et ICF International. (2012). Enquête Démographique et de Santé et à Indicateurs Multiples du Cameroun 2011. In I. International (Ed.). Calverton, Maryland, USA: ICF International

Ir, P., Korachais, C., Chheng, K., Horemans, D., Van Damme, W., & Meessen, B. (2015). Boosting facility deliveries with results-based financing: a mixed-methods evaluation of the government midwifery incentive scheme in Cambodia. *BMC Pregnancy Childbirth*, 15, 170. doi:10.1186/s12884-015-0589-x

Kandpal, Eeshani et al. Impact evaluation of performance-based financing in Nigeria. Unpublished manuscript. The World Bank.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83-119.

Ministère de la Sante Publique. (2016). Analyse Situationnelle du financement de la santé au Cameroun: Stratégie de Financement de la Santé 2017-2021. In M. d. l. S. Publique (Ed.). Yaounde, Cameroon

Paul E, Albert L, Bisala BN, *et al.* (2018) Performance-based financing in low-income and middle-income countries: isn't it time for a rethink? *BMJ Global Health* 2018;3:e000664.

RBF Health, Website accessed August 11, 2020. <https://www.rbfhealth.org/mission>

Shapira, Gil, Ina Kalisa, Jeanine Condo, James Humuza, Cathy Mugeni, Denis Nkunda, Jeanette Walldorf. (2018). "Going beyond incentivizing formal health providers: Evidence from the Rwanda Community Performance-Based Financing program." *Health Economics* 1-20.

Shen, Gordon C.; Nguyen, Ha Thi Hong; Das, Ashis; Sachingongu, Nkenda; Chansa, Collins; Qamruddin, Jumana; and Friedman, Jed. 2017 "Incentives to change: effects of performance-based financing on health workers in Zambia" *Hum Resour Health*. 2017 Feb 28;15(1):20. doi: 10.1186/s12960-017-0179-2.

United Nations Children's Fund. (2015). Enquête par grappes à indicateurs multiples (MICS5) 2014 Rapport de résultats clés. In T. U. N. C. s. Fund (Ed.), *Multiple Indicator Cluster Survey (MICS)*.

United Nations Statistical Division. (2015). Millennium Development Indicators: Country and Regional Progress Snapshots. Retrieved from <http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2015/Snapshots/CMR.pdf>

Vellez, Mercedes. (2019). "How incentive payments support Universal Health Coverage, in theory and in practice". IEG blog. <https://ieg.worldbankgroup.org/blog/how-incentive-payments-support-universal-health-coverage-theory-and-practice>. Accessed February 10, 2020.

Wagstaff, Adam, Caryn Bredenkamp, and Leander R. Buisman. 2014. Progress on Global Health Goals: Are the Poor Being Left Behind? *The World Bank Research Observer*; 29:137-162.

World Bank. (2013). Cameroon Performance-based Financing: Results from the health facility baseline survey: Results from the health facility baseline survey. In The World Bank (Ed.). Washington DC: The World Bank.

World Bank. (2016). World Development Indicators.

World Bank. (2017). Cameroon Public Expenditure Review.

World Health Organization Global Health Expenditure database. (2016). Health expenditure per capita (current US\$). (December 8). <http://data.worldbank.org/indicator/SH.XPD.PCAP>

Zeng, W., Cros, M., Wright, K. D., & Shepard, D. S. (2013). Impact of performance-based financing on primary health care services in Haiti. *Health Policy Planning*, 28(6), 596-605. doi:10.1093/heapol/czs099

APPENDIX

A1 HEALTH SERVICE VERIFICATION:

Health service verification was completed on a monthly basis by the PPA verification agents. The verification agents used the facility register and tally sheets to verify that the number of services reported by the health facility in the payment request form was consistent with the facility documentation. If the supervisor encountered any errors, these problems were corrected in the presence of the facility staff, and any fraudulent cases were tracked and documented. As an additional tool for the verification, a sample of patients for a set of health services targeted by the program was contacted either by phone or in person by local PPA staff to confirm that they received the health service reported by the health facility¹¹ and to assess patient satisfaction. The following penalties were applied for misreporting above 10% in service provision levels: 25% and 50% of the subsidy withheld for a first- and second-time offense respectively. For a third offense, the total subsidy is withheld with possible termination of contract. Additionally, the District Medical Team in collaboration with the Fund Holder Agency assessed the quality of the health services provided by PBF health facilities. This assessment used a standardized checklist to verify that a minimum quality level is met, and to calculate a quality score for the health facility ranging from 0 – 100%. The quality score was used to calculate a quality bonus that is received by the health facility.¹²

To assess the reliability of the health facility data, we examined health service counter-verification data that was collected routinely as part of the PBF program design through community client satisfaction surveys. Health service verification took place in all PBF health facilities, as well as in health facilities in the direct financing and enhanced supervision and monitoring groups. Thirty-five patients were sampled for 7 health service categories each quarter. Appendix Figure A1 shows the percentage of patients who were reported by health

¹¹ If error rates for a certain indicator surpassed 10-15% (varying slightly by region), the service was not paid and 25% of the PBF payment to the health facility was retained.

¹² An example of the quality checklist can be found here: Performance Based Financing Implementation Procedures Manual North-West Region of Cameroon North West Region, Bamenda-Cameroon: Performance Purchasing Agency; 2012. www.fbrcameroun.org/cside/contents/docs/Procedure_Manual.pdf

facilities to the PBF verification terms, who were later confirmed to have received health services at the health facility. During most quarters of the three-year study period in all three study regions, over 80% of reported patients were confirmed. The trend in confirmed patients increased slightly over time in North-West and East, with confirmation rates above 85% in all three regions during the final year of the study. Though we find reassuringly high-quality data among the treatment group health facilities, it is possible that facilities in the full control group had less incentive to keep records of all services provided. Given that we did not verify the health service data collected from the full control group, we cannot investigate this possibility. Therefore, the possibility of incomplete reporting in the full control remains a limitation of this analysis.

A2 CALCULATION OF PAYMENTS

The validated quantitative data, and the quality assessment were used to calculate performance payments for PBF health facilities. Thirty percent of the total quantitative payment was multiplied by the quality score, ranging from 0 to 100%, and was added to the quantity payment to generate the total payment amount. Quality assessments were conducted quarterly, and focused on facility management, hygiene and sanitation, as well as specific attributes of service delivery. The services delivery items included, among many other categories, listing user charges, privacy, the condition of the waiting area and consultation room, and the correct management of cases. Specifically, the quality payment worked as follows: thirty percent of the total quantitative payment was multiplied by the quality score, ranging from 0 to 100%, and was added to the quantity payment to generate the total payment amount. For example, if the health facility received a 65% quality score, and their total payment amount based on the services they provided was 597,240 CFA, the quality bonus provided to the health facility would be calculated as follows: $597,240 \times 0.30 \times 0.65 = 116,461$ CFA, and the total subsidy to the health facility would be $597,240 + 116,461 = 713,701$ CFA.

Additionally, an equity bonus was included in the calculation of performance payments. The equity bonus was paid to health facilities that faced serious structural problems making service provision more challenging. Equity Bonuses ranged from 0% to 80% of the basic subsidy. Each region applied the criteria listed below differently. North West and South West applied the same

scoring; however, few facilities in South West received the equity bonus since almost all facilities in the region were located in urban areas. The East region had a slightly difference scoring approach but also used the same criteria.

The following issues were considered in the calculation of this bonus:

- Geographical inaccessibility (hard-to-reach) that makes it difficult the retain staff;
- The size of the health area and low population densities that create viability issues (high running costs)
- Extreme poverty

The facility management committee had the authority to decide on the allocation of PBF revenue. These decisions had to be clearly documented in facility business plans.

For DF facilities, a payment equivalent in per capita terms to the PBF payments was calculated by taking the total payment amount in PBF facilities for each region was multiplied by the total population of the catchment area of a given DF facility divided by the total population served by DF facilities in the region.

A3 DIRECT OBSERVATION OF ANTENATAL CARE AND CHILD HEALTH CONSULTATIONS:

Enumerators observed a total of 575 child health consultations. For each, enumerators compared the exchange against a standardized checklist and noted whether the health worker performed the following nine routine activities: greeted the patient, washed hands, asked age, duration of the complaint, if the child is able to drink or breastfeed, if the child vomits everything, if the child is lethargic, if the child took any medicine and if the child had diarrhea. An overall quality score was calculated for each visit by calculating the proportion of these activities that was conducted. Enumerators also observed 729 ANC consultations. They compared each exchange against a standardized checklist and noted whether the health worker performed the following eleven

routine activities: took a background,¹³ asked about past issues,¹⁴ asked about current issues,¹⁵ provided iron supplementation, gave advice about warning signs,¹⁶ helped to prepare for the birth,¹⁷ checked HIV status, tested for syphilis, provided malaria prophylaxis, discussed appropriate nutrition, and checked the following vital signs: blood pressure, weight, conjunctiva, hemoglobin, rhesus, urine glucose, uterine size, fetal heartbeat and fetal presentation. As with the child health consultations, these data were used to calculate aggregate quality scores¹⁸.

¹³ A composite score ranging from 0-1 indicating whether the worker asked about the patient's age, medicines, date of last menstruation

¹⁴ A composite score ranging from 0-1 indicating whether the worker asked whether the patient had any prior deliveries, stillbirths, neonatal deaths, abortions, heavy bleeding during or after delivery, or assisted delivery.

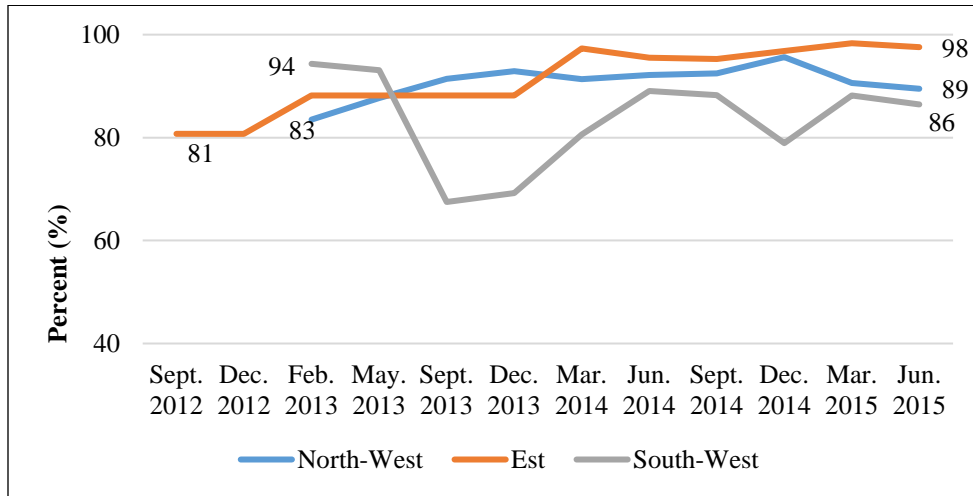
¹⁵ A composite score ranging from 0-1 indicating whether the worker asked whether the patient had any bleeding, fever, headache or blurred vision, swollen face or hands, tiredness or breathlessness, felt the baby move, or if the client noticed any other symptoms or problems related to the pregnancy.

¹⁶ A composite score ranging from 0-1 indicating whether the worker warned the patient to watch for vaginal bleeding, fever, excessive tiredness or breathlessness, swollen hands and face, and severe headache or swollen vision.

¹⁷ A composite score ranging from 0-1 indicating whether the worker advised the client to prepare for the birth, including arranging money and transportation, advised to have skilled assistance at delivery, discussed what items to have on hand, emphasized the importance of immunization and the importance of exclusive breastfeeding.

¹⁸ National protocols provide guidelines for care that is specific to the number of prior visits and gestational age of the pregnant women. As not all activities are appropriate for all consultation, the quality indices were adjusted to reflect variations by gestational age (<32 weeks, 32 to 35 weeks, and >35 weeks) and whether or not the patient was experiencing her first pregnancy.

Appendix Figure A1: percent of reported patients confirmed during verification



Appendix Table A1: PBF subsidy table

N°	Curative Care	Definition	Support documents for data collection	Unit cost in FCFA
1	Outpatient Consultations (new cases): Nurse	Number of persons consulting the health center with a new episode of illness (consulted by nurses)	Outpatient consultation register or register used for curative care consultations	200
2	Outpatient Consultations (new cases): Doctor	Number of persons consulting the health center with a new episode of illness (consulted by Medical Doctors)	Outpatient consultation register or register used for curative care consultations	650
3	Outpatient Consultations of epidemics (new cases): Doctor or nurse (free)	Number of persons consulting the health center with a new epidemic case (consulted by Medical Doctors or nurses)	Outpatient consultation register or register used for curative care consultations or special epidemics registers	1000
4	Hospital bed days (observation/Hospitalization)	Total Number of days spent by all the inpatients in the health center (for observation or awaiting referral) period limited to a maximum of 48 hours	Inpatient (hospitalization register of the health facility	400
5	Hospital bed days (observation/Hospitalization) for epidemic cases (free)	Total Number of days spent by all the inpatients epidemic cases in the health center (for observation or awaiting referral) period limited to a maximum of 48 hours	Inpatient (hospitalization register of the health facility	1500
6	Minor surgery cases	Total number of New cases of minor surgery treated in the health facility (incision of abscesses, wound sutures, circumcisions etc.)	Minor surgery register	1500
7	Referral received in the hospital	Total number of referred patients who are received at the referral hospital	Referral register of the health center, referral forms at the level of the Hospital, consultation registers of the hospital, Hospitalization registers	1500
Preventive Services/Care				
8	Children Completely Vaccinated	Children 0-11 months who received all of the following vaccines (BCG, Pentavalent 1, Pentavalent 2, Pentavalent 3 yellow fever and measles)	Vaccination register of the health facility	2500
9	VAT2 or VAT3 or VAT4 or VAT 5	Total number of women who received either VAT2 or VAT3 or VAT4 or VAT5	ANC Register and/or VAT vaccination register	1500
10	Home visits	Number of homes visited which had : appropriate collection and disposal of household refuse; a latrine in good state ; appropriate use of mosquito bed nets and use of portable water.	Home visits register signed by the Health committee representatives and the village chiefs or quarter heads	2500
11	Vitamin A supplementation (distribution)	Number of children 6 to 59 months who received Vit A	Vit A supplementation register, Vaccination Register	20
12	HIV positive Pregnant Women put on ARV prophylactic treatment	Number of HIV positive Pregnant Women put on ARV prophylactic treatment according	PMTCT Register	7000

		to the national PMTCT protocol in the month		
13	Newborn management of a baby born of an HIV positive mother.	Number of babies born of HIV positive mothers who are placed on PMTCT protocol in the month according to National directive	PMTCT Register	7000
14	Voluntary Counseling and Testing for HIV/AIDS	Number of people who came to the health facility for HIV/AIDS voluntary counseling and testing and who collected their results	VCT Register	1000
15	Cases of STIs treated	Number of new cases of STIs diagnosed and correctly treated in the month according to national protocols (Syndromic approach)	Outpatient consultation register	400
16	Cases of TB diagnosed positive by Microscopy	Number of new cases diagnosed positive by Microscopy in the health facility	TB and Lab registers	10,000
17	Cases of TB treated and healed	Total number of positive TB cases on treatment who were completely healed in the month	TB register, Lab register	20,000
Reproductive Health				
18	Normal Assisted Delivery	Total number of normal deliveries carried out by qualified (or skilled) staff (nurses) in the facility in the month	Deliveries Register (Maternity Register)	2500
19	FP: New or old acceptants on oral pills or injections	Total number of both old and new acceptants of family planning who are currently on oral pills or injections	Family Planning Register	1200
20	FP: Implants and IUD	Number of new cases of Implants and/or IUD carried out in the month	Family planning register	3000
21	Post abortive Curettage (spontaneous or induced)	Total number of new cases of curettage (post abortive) carried out in the facility in the month	Maternity and theater register	3500
22	ANC1 or ANC2 or ANC3 or ANC4	Total number of pregnant women who consulted the health facility in the month either for ANC1 or ANC2 or ANC3 or ANC4	ANC Register	500
23	IPT1 or IPT2 or IPT3	Total number of pregnant women who consulted the ANC service of the facility in the month and who took either IPT1 or IPT2 or IPT3	ANC Register	500

Note: 1 USD = 620 FCFA. The prices for each indicator were determined in concertation with the Ministry of Health taking into account cost, effort, available budget and sustainability. Relative prices between indicators also reflect public health priorities and the need to incentivize lagging indicators.

Table A2: Intervention group comparison table

	T1: PBF Complete PBF with performance bonuses for medical personnel	C1: Direct Financing PBF with subsidies that are not linked to performance	C2: Enhanced supervision and monitoring Only supervision without bonuses or autonomy	C3: Control group Status quo
Contract	Classic PBF contract	Contract stipulating the conditions of PBF for verification and supervision	Contract stipulating technical support in the form of supervision	No contract
Business plan	Yes	Yes	Simple business plan focused on intensified supervision	No business plan
Quality evaluation	Quality evaluation and feedback with quality taken into account in bonus payment	Quality evaluation with feedback as in T1, but no effect on payment	Quality evaluation with feedback as in T1	Quality evaluation with written feedback twice a year
Review/verification of service amounts	Review and verification of service quantities	Review and verification of service quantities	Review and verification of service quantities	Single quarterly statement without verification of the quantity of services produced
Payment	Payments tied to performance	Payments not tied to performance	No payment	No payment
Management autonomy	Management autonomy with control over all revenues.	Management autonomy with control over all revenues.	No management autonomy, continuation the status quo system	No management autonomy, continuation the status quo system
Monthly activity report submitted to district	Yes	Yes	Yes	Yes

Appendix Table A3: Summary of health facilities included in the impact evaluation

District	PBF	Direct financing	Enhanced supervision and monitoring	Full control	Total
Abong-Mbang	8	5	6	1	20
Doume	3	3	3	1	10
Lomie	3	2	0	4	9
Messamena	0	4	4	1	9
Nguelemendouka	3	0	0	3	6
Kette	0	2	4	3	9
Total in East	17	16	17	13	63
Kumbo East	4	8	6	4	22
Nkambe	6	5	3	4	18
Ndop	6	1	7	5	19
Fundong	4	5	4	7	20
Total in North West	20	19	20	20	79
Mamfe	2	3	3	6	14
Kumba	6	1	5	4	16
Buea	4	5	5	1	15
Limbe	4	6	5	4	19
Total in South West	16	15	18	15	64
Pilot Zone total	53	50	55	48	206

Appendix Table A4: Individual level characteristics of household members sampled at baseline*

	<i>Mean PBF</i>	<i>Mean DF</i>	<i>Mean \ESM</i>	<i>Mean control</i>	<i>Mean total</i>	<i>p-value PBF/control</i>	<i>p-value DF/control</i>	<i>p-value ESM/control</i>	<i>F-statistic</i>	<i>N</i>
<i>All household members</i>										
Age	18.41	18.76	18.00	18.35	18.37	0.846	0.197	0.245	0.155	19232
Catholic	0.44	0.37	0.32	0.36	0.37	0.140	0.858	0.416	0.150	19196
Protestant	0.36	0.40	0.42	0.43	0.40	0.099	0.498	0.701	0.379	19196
Other religion	0.15	0.15	0.13	0.14	0.14	0.582	0.701	0.886	0.875	19196
Muslim	0.05	0.08	0.13	0.07	0.09	0.437	0.696	0.114	0.143	19196
Kom	0.08	0.08	0.05	0.14	0.09	0.307	0.317	0.111	0.460	19178
Banso	0.06	0.12	0.09	0.05	0.08	0.844	0.194	0.435	0.559	19178
Other ethnicity	0.86	0.80	0.86	0.81	0.83	0.444	0.951	0.439	0.726	19178
<i>Adults > 18 years</i>										
Years of school	5.65	5.70	5.57	5.52	5.61	0.481	0.353	0.779	0.785	6807
Literacy	0.74	0.75	0.72	0.73	0.73	0.793	0.580	0.857	0.882	7991
Any school	0.88	0.87	0.87	0.85	0.87	0.299	0.636	0.651	0.775	7984
Work	0.74	0.72	0.73	0.71	0.73	0.283	0.936	0.432	0.609	7812
Agricultural work	0.60	0.57	0.58	0.58	0.58	0.704	0.858	0.953	0.945	5698
Work in retail	0.14	0.16	0.16	0.15	0.15	0.593	0.903	0.899	0.891	5698
Other type of work	0.19	0.19	0.19	0.19	0.19	0.882	0.871	0.860	0.998	7737
Never married	0.17	0.18	0.18	0.18	0.18	0.505	0.959	0.838	0.895	8038
Monogamous marriage	0.45	0.46	0.47	0.45	0.46	0.905	0.853	0.663	0.939	8038
Polygamous marriage	0.07	0.08	0.08	0.07	0.07	0.681	0.783	0.665	0.865	8038
In union	0.21	0.17	0.17	0.19	0.18	0.584	0.616	0.692	0.708	8038
Divorced or widowed	0.11	0.11	0.10	0.10	0.10	0.839	0.520	0.647	0.754	8038

*Standard errors were adjusted for facility-level clustering of observations

Appendix Table A5: Household level characteristics of households sampled at baseline*

<i>Household</i>	<i>Mean PBF</i>	<i>Mean DF</i>	<i>Mean ESM</i>	<i>Mean control</i>	<i>Mean total</i>	<i>p-value PBF/control</i>	<i>p-value DF/control</i>	<i>p-value ESM/control</i>	<i>F-statistic</i>	<i>N</i>
Total number of household members	5.61	5.47	5.57	5.64	5.57	0.895	0.328	0.696	0.769	3457
Number of women 15 - 49	1.39	1.33	1.35	1.39	1.37	0.894	0.145	0.309	0.357	3457
Number of kids under 5	1.52	1.48	1.58	1.55	1.53	0.491	0.156	0.654	0.288	3457
House with multiple flats	0.07	0.09	0.08	0.09	0.08	0.517	0.715	0.813	0.841	3457
Building with apartments	0.10	0.10	0.11	0.09	0.10	0.685	0.750	0.518	0.932	3454
Compound	0.26	0.22	0.27	0.26	0.25	0.967	0.381	0.882	0.710	3113
House	0.25	0.22	0.23	0.23	0.23	0.391	0.878	0.758	0.775	3457
Shack	0.03	0.04	0.02	0.02	0.03	0.526	0.069	0.967	0.281	3457
Other housing type	0.01	0.01	0.01	0.02	0.01	0.547	0.205	0.218	0.520	3455
Owner occupied home with mortgage	0.04	0.04	0.03	0.04	0.04	0.925	0.924	0.462	0.813	3457
Owner occupied home without mortgage	0.65	0.60	0.62	0.64	0.63	0.895	0.409	0.738	0.767	3457
Rented housing (not tied to the job)	0.14	0.16	0.15	0.15	0.15	0.750	0.807	0.980	0.957	3457
Housing rent free (other owner)	0.10	0.13	0.14	0.11	0.12	0.326	0.401	0.299	0.099	3457
Other housing payment type	0.05	0.06	0.05	0.03	0.05	0.550	0.342	0.570	0.783	3043
Piped water into the dwelling	0.03	0.02	0.03	0.03	0.03	0.978	0.482	0.966	0.689	3457
Piped water into yard/plot	0.09	0.07	0.07	0.06	0.07	0.328	0.788	0.885	0.754	3457
Piped water from public tap/standpipe	0.33	0.37	0.34	0.28	0.33	0.348	0.131	0.277	0.452	3457
Water from a well or borehole	0.06	0.09	0.14	0.07	0.09	0.718	0.567	0.055	0.146	3457
Water from a protected well	0.05	0.03	0.06	0.05	0.05	0.876	0.403	0.413	0.438	3457
Water from an unprotected well	0.06	0.05	0.06	0.05	0.05	0.546	0.829	0.443	0.865	3457
Water from a protected spring	0.08	0.09	0.07	0.09	0.08	0.742	0.966	0.565	0.915	3457
Water from an unprotected spring	0.22	0.19	0.14	0.25	0.20	0.487	0.167	0.010	0.050	3457
Surface water puddles lakes rivers	0.09	0.08	0.07	0.12	0.09	0.429	0.333	0.165	0.580	3457
Latrine pit with a slab	0.30	0.33	0.31	0.26	0.30	0.410	0.118	0.307	0.456	3455
Latrine pit without a slab	0.59	0.58	0.56	0.61	0.59	0.748	0.602	0.427	0.878	3455
Other sanitation type	0.11	0.09	0.13	0.13	0.11	0.594	0.206	0.995	0.364	3455

*Standard errors were adjusted for facility-level clustering of observations

Appendix Table A6: Baseline health service coverage

<i>Health service coverage</i>	<i>Mean PBF</i>	<i>Mean DF</i>	<i>Mean ESM</i>	<i>Mean control</i>	<i>Mean total</i>	<i>p-value PBF/control</i>	<i>p-value DF/control</i>	<i>p-value ESM/control</i>	<i>F-statistic</i>	<i>N</i>
Skilled delivery	0.77	0.75	0.78	0.76	0.77	0.864	0.920	0.790	0.981	2878
At least two ANC visits	0.86	0.90	0.91	0.87	0.88	0.604	0.216	0.140	0.164	2969
Tetanus vaccination during ANC	0.86	0.88	0.86	0.86	0.87	0.881	0.428	0.830	0.862	2971
Postnatal care	0.34	0.31	0.34	0.31	0.33	0.323	0.847	0.347	0.674	2966
Use of modern contraception	0.36	0.30	0.30	0.35	0.33	0.837	0.234	0.137	0.180	2029
Full vaccination (documented)	0.52	0.61	0.54	0.58	0.56	0.079	0.840	0.795	0.234	796
Full vaccination documented/self-report	0.53	0.66	0.59	0.62	0.60	0.355	0.593	0.495	0.497	1201
Growth monitoring	0.04	0.04	0.05	0.04	0.04	0.085	0.542	0.530	0.137	3541
Bed net use (< 5 yrs)	0.74	0.80	0.81	0.80	0.79	0.883	0.828	0.771	0.972	5786

* Standard errors were adjusted for facility-level clustering of observations

Appendix Table A7: facility level characteristics at baseline*

	<i>Mean PBF</i>	<i>Mean DF</i>	<i>Mean ESM</i>	<i>Mean control</i>	<i>Mean total</i>	<i>p- value PBF/ control</i>	<i>p- value DF/ control</i>	<i>p- value ESM/ control</i>	<i>F- statisti c</i>	<i>N</i>
<i>Facility characteristics</i>										
Number of beds in the health facility	8.07	9.98	11.70	9.26	9.84	0.453	0.788	0.226	0.244	185
Electricity in the health facility	0.70	0.78	0.69	0.77	0.73	0.410	0.914	0.363	0.626	206
Piped water in the health facility	0.40	0.38	0.35	0.35	0.37	0.665	0.792	0.927	0.947	206
Facility has an incinerator	0.08	0.22	0.24	0.23	0.19	0.033	0.914	0.932	0.027	206
Latrine in the health facility	0.85	0.84	0.85	0.79	0.83	0.457	0.540	0.409	0.853	206
Facility open 24 hours	0.66	0.72	0.64	0.71	0.68	0.607	0.899	0.439	0.775	206
Water towel and soap in exam room	0.46	0.43	0.47	0.45	0.45	0.897	0.858	0.805	0.976	199
Secure box for sharps	0.80	0.86	0.80	0.83	0.82	0.708	0.715	0.668	0.832	200
User fees for consultation posted	0.38	0.32	0.36	0.35	0.35	0.810	0.723	0.921	0.939	206
User fees for laboratory services posted	0.34	0.35	0.37	0.23	0.32	0.257	0.225	0.154	0.460	195
Child weighing scale	0.87	0.88	0.94	0.83	0.88	0.595	0.460	0.069	0.222	202
Height Measure	0.41	0.43	0.45	0.53	0.46	0.253	0.303	0.430	0.663	191
Tape measure	0.96	0.98	1.00	0.96	0.98	0.903	0.539	0.153	0.167	204
Blood pressure instrument	0.86	0.90	0.87	0.85	0.87	0.836	0.468	0.777	0.895	199
Thermometer	0.98	0.94	0.95	1.00	0.97	0.317	0.079	0.079	0.067	204
Stethoscope	0.96	0.92	0.91	0.91	0.93	0.353	0.951	0.918	0.630	202
Lab services	0.74	0.80	0.82	0.77	0.78	0.686	0.727	0.557	0.760	206
Blood test	0.34	0.42	0.48	0.54	0.45	0.084	0.314	0.574	0.347	159
Malaria test	0.97	1.00	0.91	0.97	0.96	0.970	0.317	0.223	0.101	160
TB test	0.13	0.28	0.20	0.19	0.20	0.501	0.375	0.864	0.461	159
HIV test	0.11	0.23	0.18	0.22	0.18	0.194	0.880	0.703	0.402	158
Facility provided immunization	0.98	0.96	0.95	0.98	0.97	0.944	0.582	0.366	0.734	206
Facility provides antenatal care	0.98	0.98	0.98	1.00	0.99	0.318	0.317	0.318	0.392	206

Appendix Table A7 continued: Facility level characteristics at baseline*

	<i>Mean PBF</i>	<i>Mean DF</i>	<i>Mean ESM</i>	<i>Mean control</i>	<i>Mean total</i>	<i>p- value PBF/ control</i>	<i>p- value DF/ control</i>	<i>p- value ESM/ control</i>	<i>F- statisti c</i>	<i>N</i>
<i>Baseline health service provision</i>										
Antenatal care	23.29	14.52	17.86	20.57	19.05	0.691	0.270	0.643	0.479	1110
Family planning	2.82	1.40	3.25	3.02	2.62	0.876	0.159	0.870	0.072	1110
Delivery	7.57	4.71	6.43	7.76	6.61	0.936	0.139	0.563	0.320	1097
Postpartum care	8.81	5.53	7.61	10.22	7.98	0.639	0.092	0.378	0.147	1110
Third polio vaccine	21.39	14.77	20.06	23.90	19.92	0.668	0.051	0.468	0.163	1110
Meningitis vaccine	28.07	22.18	35.48	46.65	32.61	0.174	0.028	0.398	0.133	1110
Measles vaccine	19.80	13.65	18.89	20.90	18.24	0.847	0.124	0.688	0.265	1110
Tetanus vaccine during pregnancy	32.96	20.77	30.43	32.84	29.18	0.990	0.100	0.764	0.246	1110

*Standard errors adjusted for facility-level clustering of observations

Appendix Table A8: Operational financing available at the health facility and PBF payments assessed at endline in the performance-based financing and direct financing groups

Non-standardized financing			
<i>Total (cost recovery, Ministry of health, subsidies)</i>			
	PBF	DF (direct financing)	p-value
Q1-2014 received	3,420,226 (2,079,623, 4,760,829)	4,163,127 (2,174,893, 6,151,360)	0.5281
Q2-2014 received	3,339,472 (1,990,924, 4,688,020)	4,044,135 (2,326,276, 5,761,994)	0.5143
Q3-2014 received	3,799,585 (1,700,043, 5,899,126)	4,395,996 (1,246,833, 7,545,158)	0.748
Q4-2014 received	3,876,873 (1,636,670, 6,117,075)	3,873,040 (2,073,397, 5,672,683)	0.9979
n	53	48	101
<i>Subsidies</i>			
Q1-2014 received	1,322,834 (943,464, 1,702,204)	1,725,858 (1,214,379, 2,237,337)	0.2009
Q2-2014 received	1,241,536 (757,133, 1,725,940)	1,725,858 (1,214,379, 2,237,337)	0.1805
Q3-2014 received	1,757,934 (656,930, 2,858,938)	1,725,858 (1,214,379, 2,237,337)	0.5654
Q4-2014 received	1,428,642 (351,293, 2,505,990)	1,264,029 (723,122, 1,804,936)	0.7897
n	53	48	101
<i>Difference between total financing and subsidies</i>			
	PBF	DF (direct financing)	p-value
Q1-2014 received	2,097,392 (951,186, 3,243,599)	2,437,269 (518,605, 4,355,933)	0.7554
Q2-2014 received	2,097,936 (902,990, 3,292,881)	2,437,269 (518,605, 4,355,933)	0.8437
Q3-2014 received	2,041,651 (788,702, 3,294,599)	2,437,269 (518,605, 4,355,933)	0.5582
Q4-2014 received	2,041,651 (788,702, 3,294,599)	2,437,269 (518,605, 4,355,933)	0.8855
n	53	48	101
Financing standardized by number of health workers			
<i>Total standardized by number of health workers</i>			
	PBF	DF (direct financing)	p-value
Q1-2014 received	559,007 (425,894, 692,119)	587,480 (406,054, 768,906)	0.7962
Q2-2014 received	442,595 (348,302, 536,888)	550,209 (395,925, 704,493)	0.2212
Q3-2014 received	436,624 (326,889, 546,358)	471,578 (331,729, 611,428)	0.6895
Q4-2014 received	467,895 (303,072, 632,719)	528,527 (379,769, 677,286)	0.5869
n	49	43	92
<i>Subsidies standardized by number of health workers</i>			
Q1-2014 received	295,233 (212,497, 377,968)	391,902 (220,262, 563,541)	0.2906
Q2-2014 received	242,738 (168,911, 316,564)	361,116 (221,024, 501,208)	0.1223
Q3-2014 received	238,237 (164,959, 311,514)	260,400 (167,355, 353,444)	0.7039
Q4-2014 received	212,286 (116,566, 308,005)	312,874 (178,213, 447,535)	0.2159
n	49	43	92
<i>Difference between total financing and subsidies standardized by the number of health workers</i>			
Q1-2014 received	263,774 (187,366, 340,182)	195,578 (125,382, 265,774)	0.1934
Q2-2014 received	199,858 (142,919, 256,797)	189,093 (117,579, 260,608)	0.8110
Q3-2014 received	198,387 (141,423, 255,351)	211,178 (110,152, 312,205)	0.8191
Q4-2014 received	255,610 (136,328, 374,892)	215,653 (137,784, 293,523)	0.5831
n	49	43	92

Appendix Table A9: Satisfaction with antenatal care consultations reported during facility exit interviews†

Panel A	(1) Overall satisfaction score	(2) Reasonable registration fees	(3) Reasonable lab fees	(4) Reasonable medicine fees	(5) No additional payment	(6) Clean health facility	(7) Reasonable wait time
Post indicator	0.006 [0.034]	0.055 [0.097]	0.074 [0.075]	0.088 [0.090]	0.010 [0.060]	-0.045 [0.075]	-0.009 [0.079]
PBF/Post interact	0.086* [0.048]	0.037 [0.128]	0.154 [0.113]	0.190 [0.134]	-0.043 [0.078]	0.241** [0.111]	0.161 [0.115]
DF/Post interact	0.051 [0.044]	-0.051 [0.142]	-0.051 [0.129]	-0.069 [0.121]	0.067 [0.092]	0.228** [0.106]	0.014 [0.129]
ESM/Post interact	-0.019 [0.049]	-0.085 [0.127]	-0.027 [0.127]	-0.127 [0.118]	-0.020 [0.087]	0.002 [0.111]	-0.055 [0.134]
p-PBF vs. DF	0.419	0.523	0.129	0.044	0.230	0.903	0.269
p-PBF vs. ESM	0.036	0.309	0.185	0.015	0.776	0.040	0.127
p-PBF vs. control	0.077	0.774	0.176	0.158	0.586	0.032	0.163
Baseline mean control	0.853	0.804	0.782	0.754	0.885	0.787	0.738
N	730	669	665	689	723	730	727
Panel B	(8) Enough privacy during visit	(9) Adequate hours	(10) Courteous health staff	(11) Good health worker communicat ion	(12) Sufficient visit time with health worker	(13) Easy to get prescribed medicines	
Post indicator	-0.025 [0.061]	-0.026 [0.041]	-0.041 [0.036]	-0.031 [0.024]	0.055 [0.052]	0.018 [0.068]	
PBF/Post interact	0.042 [0.086]	0.154** [0.071]	0.037 [0.063]	0.106** [0.050]	-0.045 [0.074]	0.000 [0.079]	
DF/Post interact	0.149 [0.093]	0.032 [0.055]	0.070 [0.053]	0.039 [0.052]	0.139* [0.080]	0.041 [0.085]	
ESM/Post interact	-0.010 [0.098]	-0.079 [0.062]	0.019 [0.051]	0.101 [0.077]	0.030 [0.101]	-0.040 [0.075]	
p-PBF vs. DF	0.254	0.073	0.599	0.308	0.026	0.488	
p-PBF vs. ESM	0.605	0.002	0.783	0.961	0.443	0.402	
p-PBF vs. control	0.629	0.033	0.556	0.038	0.544	0.997	
Baseline mean control	0.902	0.900	0.967	0.951	0.852	0.883	
N	728	724	730	725	728	716	

* = $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

† Results from difference-in-differences regression models examining the effect of PBF on satisfaction with antenatal care components reported by patients during facility exit interviews. Regression models adjusted for individual (age, literacy, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level. For each question, a statement was read, and women were asked if they agreed, were neutral, or they disagreed. Binary variables were created by coding responses as “1” if a woman agreed, and “0” otherwise. Overall satisfaction scores were calculated by averaging over these twelve components. An overall score of “1” indicates that a woman agreed with all twelve questions, while a score of “0” indicates that she either disagreed or was neutral on all twelve questions.

Appendix Table A10: Satisfaction with child health consultations reported during facility exit interviews†

Panel A	(1) Overall satisfaction score	(2) Reasonable registration fees	(3) Reasonable lab fees	(4) Reasonable medicine fees	(5) No additional payment	(6) Clean health facility	(7) Reasonable wait time
Post indicator	-0.036 [0.026]	-0.054 [0.061]	-0.143 [0.118]	-0.045 [0.067]	-0.004 [0.055]	-0.141 [0.099]	-0.111** [0.055]
PBF/Post interact	0.099*** [0.037]	0.112 [0.101]	0.347** [0.175]	0.043 [0.124]	-0.007 [0.117]	0.227* [0.133]	0.110 [0.097]
DF/Post interact	0.054 [0.040]	0.019 [0.076]	0.331* [0.167]	0.111 [0.103]	-0.020 [0.077]	0.136 [0.118]	0.143 [0.097]
ESM/Post interact	0.022 [0.045]	0.074 [0.080]	0.420** [0.166]	0.033 [0.120]	-0.055 [0.090]	-0.049 [0.131]	0.021 [0.100]
p-PBF vs. DF	0.280	0.325	0.925	0.595	0.908	0.403	0.777
p-PBF vs. ESM	0.092	0.709	0.685	0.945	0.696	0.019	0.442
p-PBF vs. control	0.009	0.268	0.050	0.731	0.953	0.090	0.259
Baseline mean control	0.881	0.957	0.846	0.854	0.904	0.868	0.943
N	614	488	369	544	605	612	608
Panel B	(8) Enough privacy during visit	(9) Adequate hours	(10) Courteous health staff	(11) Good health worker communicat ion	(12) Sufficient visit time with health worker	(13) Easy to get prescribed medicines	
Post indicator	-0.098 [0.098]	-0.046 [0.052]	0.062 [0.062]	0.017 [0.073]	0.091 [0.066]	-0.002 [0.059]	
PBF/Post interact	0.336*** [0.124]	0.085 [0.068]	-0.012 [0.079]	0.053 [0.092]	-0.094 [0.087]	0.055 [0.082]	
DF/Post interact	0.202 [0.131]	0.036 [0.068]	-0.080 [0.077]	-0.080 [0.105]	-0.018 [0.111]	0.068 [0.087]	
ESM/Post interact	0.093 [0.115]	-0.116 [0.081]	-0.101 [0.082]	0.055 [0.103]	-0.031 [0.104]	0.112 [0.106]	
p-PBF vs. DF	0.256	0.427	0.323	0.170	0.461	0.890	
p-PBF vs. ESM	0.021	0.007	0.214	0.988	0.507	0.594	
p-PBF vs. control	0.007	0.210	0.876	0.566	0.279	0.498	
Baseline mean control	0.774	0.942	0.887	0.830	0.830	0.925	
N	612	608	613	606	609	610	

* = $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ † Results from difference-in-differences regression models examining the effect of PBF on satisfaction with child health consultations reported by mothers during facility exit interviews. Regression models adjusted for individual (child age, child sex, maternal literacy, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level. For each question, a statement was read, and women were asked if they agreed, were neutral, or they disagreed. Binary variables were created by coding responses as “1” if a woman agreed, and “0” otherwise. Overall satisfaction scores were calculated by averaging over these twelve components. An overall score of “1” indicates that a woman agreed with all twelve questions, while a score of “0” indicates that she either disagreed or was neutral on all twelve questions.

A4 REGRESSION SPECIFICATIONS AND ANALYSIS OF HOUSEHOLD CARE SEEKING BEHAVIOR

The main regression specification used in the analysis of facility-level data is presented below¹⁹:

$$Y_{jt} = \alpha_j + \beta_1 I_{2015} + \beta_2 PBF I_{2015} + \beta_3 DF I_{2015} + \beta_4 ESM I_{2015} + \beta' X_j + \beta' X_{jt} + \varepsilon_{jt} \quad (1)$$

Where Y_{jt} is receipt of the health service Y in enumeration area j in survey year t . α_j is an health facility fixed effect, I_{2015} is an indicator variable that is equal to 0 in 2012 (baseline) and 1 in 2015 (endline). $\beta_1 PBF I_{2015}$, $\beta_2 DF I_{2015}$, and $\beta_3 ESM I_{2015}$ are interaction terms between the *PBF*, *DF*, and *ESM* groups and the post indicator, respectively. Analysis at the facility level included the following controls: type of health facility (public/religious/private) and location of the health facility (urban/rural). Additionally, when the analysis was at the individual level (i.e. women sampled from ANC visits, care givers sampled from child health consultations) the following individual level controls were also included: age, sex, marital status, and education level. Standard errors were clustered at the health facility level.

$$Y_{ijt} = \alpha_j + \gamma_{2015} + \beta_1 PBF I_{2015} + \beta_2 DF I_{2015} + \beta_3 ESM I_{2015} + \beta' X_{it} + \varepsilon_{ijt} \quad (2)$$

Where Y_{ijt} is receipt of the health service for woman/pregnancy i in enumeration area j in survey year t . α_j is an enumeration area fixed effect, γ_{2015} is an indicator variable that is equal to 0 in 2012 (baseline) and 1 in 2015 (endline). The treatment variable is based on the assigned catchment area where the household is located; These interaction terms measure the treatment effect in each group and can be interpreted as the difference in the change in health service use over the study period between the control group, and each treatment group respectively. $\beta' X_{it}$ is a vector of control variables at the individual level (age, marital status, education level, religion, ethnicity, working status and type of work), and at the household level (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Control variables were included in all analysis of household level data. Standard errors were clustered at the level of the enumeration area.

For the household survey, a random sample of 16 to 20 households was selected in each health facility catchment area. The analysis of the household survey in this report starts from the assumption that household members seek care in the health facility closest to where they live, or in other words that people living in the catchment area of a facility obtain health care in that facility.

However, it is apparent from the baseline survey data that households do not always seek care from the closest health facility in their health zone (Appendix table A11). The household survey analysis assigns a treatment group (*PBF*, *DF*, *ESM* or *C*), to each household in the study, which

¹⁹ A similar specification was used in most facility level analysis; however, in all analyses involving direct observation or exit interview data we used an alternative specification without facility fixed effects. Therefore, this analysis instead included treatment group indicator variables to control for baseline differences between groups. For consistency, all analysis of women sampled from antenatal care, and of caregivers sampled from child health consultations used this alternative specification. Moreover, sampling for these services was limited in many health facilities because antenatal care is only provided on certain days or the week, and due to low patient flows in smaller facilities. For this reason, analysis of patients sampled from health facilities was not restricted to health facilities represented in both the baseline and endline survey data.

represents the treatment assignment of the closest health facility to the sampled household. Additionally, both the baseline and the endline household survey included information about the name of the health facility where the household sought reproductive health care. Using this information, we created a variable that measures whether women sought care in a health facility consistent with their assigned treatment group, a health facility assigned to another treatment group, a non-randomized hospital, or a health facility outside of the study area. At baseline, for antenatal care, for example, 44.8% of women sought care in a health facility assigned to their own treatment group, but 11.1% sought care in a higher level hospital (not included in the randomization conducted for the impact evaluation) and 22.2% sought care in other health facilities beyond their own health zone (18.7% in other facilities assigned to other treatment groups in our study sample and 3.2% in facilities not included in the study sample). Another 7.8% did not seek any antenatal care and we are missing information about the service location for the remaining 15.7%. If we focus on women for whom we have information about the service location, 52.3% sought care in in their “assigned” treatment group, and if we further exclude women who did not seek any antenatal care, this percentage increases to 57.6%. This “health care shopping” behavior whereby households bypass the closest health facility is also present for deliveries and postnatal care: at baseline, focusing on women for whom we have information about the service location and who sought care in a facility, only 51.9% delivered, and only 56.1% sought postnatal care in a facility with their corresponding treatment group. At endline, those percentages are slightly higher, but not very substantially: focusing again on women for whom we have information about the service location and who sought care in a facility, 60.9% obtained antenatal care, 55% delivered, and 60.6% sought postnatal care in a facility assigned to their treatment group.

When the household is indeed seeking care in a health facility that is consistent with their assigned treatment group, this assignment is correctly done. However, when the household seeks care in another facility, this assignment between household and study group is potentially erroneous, leading to measurement error. This measurement error would introduce statistical noise in the analysis and reduce our capacity to measure potential impacts of the interventions (attenuation bias). Another interpretation of these patterns is to see it as non-compliance with assigned treatments. The intent-to-treat (ITT) model estimated remains valid. The causal estimates the ITT model creates may however not fully capture the causal effect of the treatment relative to a “clean” control, but rather measure the causal impact of having a treated facility closer compared to people living further away from a treated facility. These estimates are likely below the true causal effect of the intervention. This is a substantial limitation of the household survey analysis that needs to be kept in mind.

The statistical analysis and interpretation of the household survey would be further complicated if this health care “shopping” behavior was driven or reinforced by the introduction of PBF or the interventions implemented in the direct financing and enhanced supervision and monitoring groups. If this was the case, this could introduce a bias in the estimates going further than the attenuation bias described above. Appendix Tables A12 to A14 investigate whether the implementation of PBF or the other interventions have directly influenced household’s health care seeking behaviors. For antenatal care, deliveries and postnatal care, they report results from multinomial logit difference-in-differences regression models where the four options for the household are: not seeking care, seeking care at the assigned facility, i.e. the closest one in the health zone, seeking care in an unassigned facility, i.e. another facility of the same level potentially randomized into a different impact evaluation group, and seeking care at a non-

randomized facility, generally a higher level hospital which was not included in the randomization.

Overall, the results do not suggest that the health care seeking behavior is driven or even significantly influenced by the introduction of PBF or the other interventions (direct financing and enhanced monitoring and supervision). We also conducted this analysis using ordinary least squares with a binary outcome equal to 1 if the respondent sought care in a health facility in their assigned treatment group, and zero otherwise. We found no evidence that health care shopping was affected by PBF in this analysis, consistent with the results from the multinomial logit²⁰. Health care shopping behavior by households was widespread in Cameroon at baseline in 2012 and continues to be widespread at endline in 2015 but does not appear to be a consequence of the introduction of PBF. We therefore decided to keep the presentation of the household survey results in which the analysis assumes that household members seek care in the health facility closest to where they live.

²⁰ Results available on request

Appendix table A11: health care seeking behavior†

Baseline	Antenatal care		Delivery care		Postnatal care	
	N	%	N	%	N	%
Did not receive the health service	214	7.78	585	21.26	1,825	66.34
Received the health service in assigned treatment group	1,231	44.75	951	34.57	440	15.99
Received the health service in different treatment group	514	18.68	361	13.12	154	5.6
Received the health service in a facility outside the study area	88	3.2	72	2.62	44	1.6
Received the health service in a non-randomized hospital	304	11.05	304	11.05	138	5.02
Missing data on service location	398	15.73	332	19.03	141	15.38
N	2,751		2,751		2,751	
Endline	Antenatal care		Delivery care		Postnatal care	
	N	%	N	%	N	%
Did not receive the health service	226	6.82	582	17.57	1,954	59
Received the health service in assigned treatment group	1,626	49.09	1,358	41	705	21.29
Received the health service in different treatment group	435	13.13	395	11.93	177	5.34
Received the health service in a facility outside the study area	315	9.51	200	6.04	102	3.08
Received the health service in a non-randomized hospital	292	8.82	321	9.69	171	5.16
Missing data on service location	414	13.43	261	12.58	194	14.38
N	3,312		3,312		3,312	

†Percentages calculated from household survey data among sampled women who had been pregnant in the 24 months before the survey.

Appendix Table A12: Health care shopping for antenatal care†

	(1) No ANC	(2) ANC in assigned treatment group facility	(3) ANC in unassigned treatment group facility	(4) ANC in non- randomized facility
Post indicator	0.001 [0.014]	0.040 [0.045]	-0.060* [0.033]	-0.014 [0.018]
PBF/Post interact	0.012 [0.018]	0.065 [0.055]	-0.021 [0.039]	-0.019 [0.024]
DF/Post interact	-0.004 [0.018]	0.016 [0.054]	0.045 [0.038]	-0.015 [0.023]
ESM/Post interact	0.015 [0.020]	-0.017 [0.057]	0.005 [0.041]	0.016 [0.023]
N	5407	5407	5407	5407

* = $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

† Results from multinomial logistic difference-in-differences regression models examining the effect of PBF on facility bypassing for reproductive health care. Regression model adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work), household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation) and facility-level control variables at baseline (type of health facility, urban/rural). Standard errors were clustered at the health facility level.

Appendix Table A13: Health care shopping for skilled delivery†

	(1)	(2)	(3)	(4)
	No skilled delivery	Skilled delivery in assigned treatment group facility	Skilled delivery in unassigned treatment group facility	Skilled delivery in non-randomized facility
Post indicator	-0.036* [0.020]	0.064 [0.040]	-0.060** [0.030]	-0.001 [0.017]
PBF/Post interact	0.018 [0.028]	0.038 [0.050]	0.042 [0.033]	-0.038 [0.025]
DF/Post interact	-0.000 [0.030]	0.003 [0.050]	0.072** [0.034]	-0.012 [0.023]
ESM/Post interact	0.016 [0.028]	0.004 [0.048]	0.039 [0.037]	-0.018 [0.022]
N	5419	5419	5419	5419

* = $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

† Results from multinomial logistic difference-in-differences regression models examining the effect of PBF on facility bypassing for reproductive health care. Regression model adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work), household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation) and facility-level control variables at baseline (type of health facility, urban/rural). Standard errors were clustered at the health facility level.

Appendix Table A14: Health care shopping for postnatal care†

	(1)	(2)	(3)	(4)
	No postnatal care	Postnatal care in assigned treatment group facility	Postnatal care in unassigned treatment group facility	Postnatal care in non-randomized facility
Post indicator	-0.059** [0.025]	0.047* [0.027]	-0.031** [0.016]	0.005 [0.010]
PBF/Post interact	-0.030 [0.032]	0.056* [0.033]	0.030* [0.018]	-0.018 [0.014]
DF/Post interact	0.013 [0.030]	-0.015 [0.031]	0.041** [0.019]	-0.006 [0.012]
ESM/Post interact	0.014 [0.031]	0.010 [0.032]	0.014 [0.019]	-0.006 [0.013]
N	5634	5634	5634	5634

* = $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

† Results from multinomial logistic difference-in-differences regression models examining the effect of PBF on facility bypassing for reproductive health care. Regression model adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work), household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation) and facility-level control variables at baseline (type of health facility, urban/rural). Standard errors were clustered at the health facility level.

